

## Illinois State University ISU ReD: Research and eData

---

Faculty and Staff Publications – Milner Library

Milner Library

---

3-2010

# Digital Media Reviews: Themefinder [Review]

Anne Shelley

*Illinois State University*, [aeshell@ilstu.edu](mailto:aeshell@ilstu.edu)

Follow this and additional works at: <http://ir.library.illinoisstate.edu/fpml>



Part of the [Library and Information Science Commons](#)

---

### Recommended Citation

Shelley, Anne, "Digital Media Reviews: Themefinder [Review]" (2010). *Faculty and Staff Publications – Milner Library*. Paper 7.  
<http://ir.library.illinoisstate.edu/fpml/7>

This Article is brought to you for free and open access by the Milner Library at ISU ReD: Research and eData. It has been accepted for inclusion in Faculty and Staff Publications – Milner Library by an authorized administrator of ISU ReD: Research and eData. For more information, please contact [ISUReD@ilstu.edu](mailto:ISUReD@ilstu.edu).

split between two databases sold separately could be a stumbling block. Still, it's such an obviously good idea that someone should be able to solve this minor problem.

### Conclusion

A common characteristic of all audio databases, certainly the ones mentioned here—*CWM*, *Smithsonian Global Sound*, and *Naxos Music Library*—is that they depend on the licensing of a limited number of record labels that have very selectively issued a tiny portion of the music available in the world. Many instructors of specific topics in ethnomusicology nowadays are just as likely to use their own field recordings and Youtube as they are commercial recordings, and so a database like *CWM* would be mostly supplemental in these cases. For broader “world

music survey” courses, as well as for non-music courses needing illustrations of music's ability to “connect on the feeling level with people and their conditions,” it can certainly play a more central role insofar as it continues to offer a representative sampling of music in the world, particularly of those many currents of “interchange” occurring within the last thirty or forty years. Amid recent discussions of the decline of the CD as a medium, and the resulting difficulties for libraries in acquiring music that may only be available for individual users to download, content-hosting databases such as those from Alexander Street Press and Naxos may be the most hopeful long-term solution.

ALEC McLANE

Wesleyan University

**Themefinder.** Center for Computer Assisted Research in the Humanities, Stanford University and the Cognitive and Systematic Musicology Laboratory, Ohio State University. <http://www.themefinder.org> and <http://rism.themefinder.org> (Accessed August–September 2009). [Requires a Web browser and an Internet connection.]

Themefinder was developed in the mid-to-late-1990s as a collaborative project between Stanford University's Center for Computer Assisted Research in the Humanities and the Ohio State University's Cognitive and Systematic Musicology Laboratory. Themefinder allows users to search for short musical incipits, or themes. If a user knows a melody's or musical work's meter, key, mode, pitch names, or general or specific intervals, he or she may input such data to discover that work's title, composer, or genre. This type of searching is one cornerstone of the multi-disciplinary field of Music Information Retrieval (MIR), which brings together the expertise of musicologists, music theorists, computer scientists, library and information science specialists, and audio engineers. Themefinder moderators encourage use of the musical themes in teaching, non-commercial research, and personal instances, but the site does note that both the graphical notation and the underlying data are protected by international copyright laws. There is no charge to use Themefinder.

Themefinder provides a Web-based interface to Humdrum, a complex software sys-

tem that allows users to pose and answer music-related research questions. Using Humdrum's syntax, Themefinder can search one of four databases containing musical themes or incipits: Classical Instrumental Music, International Folksongs, Latin Motets of the Sixteenth Century, and RISM, series AII. Encoding of Themefinder's original collection of Classical themes began in 1996; the Classical database covers the broadest scope of the three, from Bach to Shostakovich, and contains around 10,000 entries. Currently, there are around 7,000 themes in the Essen Folksong Collection and nearly 20,000 themes in the Latin Motet index. RISM incipits have been available through Themefinder since early 2009 and number over 55,000.

### Content and Scope

All incipits in the Classical and Folksong databases have at least three—most typically, four—measures of the beginning of a work or movement. The Renaissance database does not include bar lines, but all themes in this database display at least the opening phrase of each vocal part. The instrumentation of themes in each database

differs greatly. Each Renaissance theme represents each vocal part of a motet separately. For example, Byrd's *Emendemus in melius no. 4* shows five incipits, three in treble clef and two in bass clef, each with a voice designation in the record (one of five, three of five, etc.). The Essen Folksong database contains a combination of vocal and instrumental themes, while the Classical database contains only instrumental themes. Although the Classical database contains multi-instrumental works, it does not provide a separate incipit for each instrument associated with a single work. So the record for the fifth movement of Purcell's Sonata No. 3 in A minor for two violins and harpsichord displays a single theme for all three instruments. Indications of instrumentation are not provided.

A user may click on each theme to view a structured metadata record that typically includes the composer's name along with the work's title, genre, and instrumentation; this is true for any of the three databases. In the case of the Renaissance motets, a publisher's catalog number, original publisher, and date and location of first publication are also included. Unique metadata in the Essen Folksong database includes an alphanumeric original document designation, a catalog number, a metric classification, and a genre designation in the composition's original language. Though the scope is truly global, a majority of the works represented in the Folksong database are from areas in and around Germany and China. A typical record for a theme stored in the Classical database includes information on the theme's composer, as well as the theme's title and opus and movement numbers.

Each theme record displays a series of icons below the staff, and each of these icons is linked to supplemental materials or information related to that theme. One icon, "i," leads to the basic biographical and work metadata described above, while an eighth-note icon links to a MIDI file of the theme, played in a separate browser screen in QuickTime. A "+" icon caches that particular theme to a clipboard within Themefinder; a user may add multiple themes to the clipboard within a Web session. Some themes have "H" icons that display the underlying Humdrum code. Finally, there are separate icons—"C," "W,"

or "M"—that lead the user to lists of themes within a certain collection, work, or movement, respectively. For instance, a grouping of collection themes from Book I of Bach's Well-tempered Clavier includes each Prelude and Fugue in Book I. A list of work themes for Book I, Prelude No. 4 would show just that Prelude and Fugue No. 4.

### Applications

Pedagogical applications of Themefinder will not be explored here in detail, but it is worth highlighting the Haydn/Mozart String Quartet Quiz as one manifestation of Themefinder's capabilities beyond naming "that tune." After providing some personal information regarding his or her musical background and training, a user may listen to a MIDI file of an extended theme taken from either a Mozart or Haydn string quartet. Then, the user answers a short series of questions, including his or her level of certainty of the composer, whether or not he or she has heard the movement before and, if so, the user enters the name and movement of the quartet. Once the user has answered, statistics are displayed for correct answers by day, week, and cumulatively. There is also a rating scale so one can determine one's level of expertise relative to the norm.

A link on the Themefinder homepage directs the user to a thread of "new" links, a presumably outdated page as the most recent contribution is from mid-2000. This page lists user-contributed "relevant links to themes in the Themefinder database." Once a user has searched for and located a theme within Themefinder, he or she has the option of linking that theme to a related, significant external resource. The user is prompted to enter title, author, and URL information, as well as a description of the site. From this information, Themefinder provides brief annotations for the external links, which connect to a range of resources, including program notes, MIDI files, and articles related to particular themes. These annotations are provided both on the individual theme's page and on the "new links" page. Unfortunately, there are some maintenance and consistency issues with this facet of Themefinder. The majority of the external links are broken and while most entries on the "new

links" page reliably link back to their respective themes, there is not always a corresponding link from the theme page to the external resource. If kept up-to-date and monitored for accuracy, relevancy, and currency by the webmaster, this aggregation of resources could certainly provide a meaningful context for Themefinder's content. In addition, providing a niche group of users with an opportunity to contribute content may increase activity on the site. In its state at the time of this writing, however, this seemingly unmoderated option serves only to blemish Themefinder's credibility as an academic resource. Perhaps the fact that this supplemental page has been so neglected highlights Themefinder's primary purpose as a notation search engine and not as a more traditional bibliographic tool.

### Interface and Searching

In general, navigation across Themefinder's subpages is difficult. There are no persistent headers, menus, or mappings that appear on every page of the site, which forces the user to primarily use browser navigation options. However idiosyncratic a page header might be, though, Themefinder does make an effort to hyperlink each header back to the homepage and main search interface. Some examples of these headers include "Theme Information" for pages with a specific theme's content and "Themefinder Results," which displays a listing of themes as search results. Themefinder appears to display equally well across the browsers Internet Explorer 8.0, Mozilla Firefox 5.0, and Safari 4.0. Overall, Themefinder also performed well in Safari on the iPod Touch, but MIDI files would not play in this case.

One cannot simply browse all themes within a database, so the user must enter at least one search criterion. The three databases may be searched separately or all together. Themefinder's main page displays its primary search interface, which allows the user to search by pitch, interval, scale degree, gross contour (general melodic shape), or refined contour (specific melodic shape). Each option may be searched independently or in combination with others. In the same search interface, the user also has a few options to refine his or her search. Though the default search for melodic segments is at the beginning of

a theme, the user may also search for a string at any point in the theme. One can also search a specific key within the chromatic scale, by major or minor mode, and by meter. Each search field displays a brief annotation to assist the user in correctly entering machine-parsable data. For instance, instructions adjacent to the "pitch" search field indicate that a user may enter the letters A-G, the pound symbol (#) for a sharp sign and the minus sign (-) for a flat sign. Expanded guidelines include wildcard search options for an unknown pitch, a single optional note or several of them, any number of repeated notes, and arbitrary accidentals.

In addition to the standard search options, the Essen Folksong collection, the collection of sixteenth-century Latin motets, and themes included from RISM series AII all have unique search interfaces that allow the user to perform text-only queries. Each of these separate interfaces, however, exhibits oddities that are both frustrating and confusing. The Folksong collection offers pre-populated dropdown lists for a song's genre and its country of origin, however, this search method may not be the most reliable one to use, as a search for songs from China yields zero results even though records for Chinese folksongs have displayed as results in various searches. Even more puzzling is the fact that Themefinder indicates in the dropdown menu, alongside the name of the country, that there are 1,972 themes representing Chinese folksongs. Similarly, when one attempts to browse themes by country, the number of themes in the results set is nowhere near the number indicated in the dropdown menu. For instance, Themefinder claims there are 105 songs of Swiss origin, yet only eight results are displayed. Search results from this database also differ substantially depending on whether one's search is performed using Themefinder's primary interface or the folksong-specific page. A search for the pitches "CA" on Themefinder's main page retrieves 143 records, while only two results display in a comparable search from the collection-specific page. When searching the Latin Motet database, there are no glaring malfunctions in the search options specific to that collection—RISM number, Publisher name and location, to name a few—but performing any of Themefinder's standard

searches by pitch, interval, or contour yields no results whatsoever. There is a free-text search box for song titles that appears to function normally for both the Folksong and Latin Motet collections. Strangely, the Classical database does not allow for free-text searching.

Searching by pitch names is not case-sensitive. However, users will need to take case into consideration when searching by interval—"m" represents "minor," "M" searches for "major"—and by refined contour, where a lowercase "u" indicates an upward step and an uppercase "U" denotes an upward leap in the melodic line. The pitch search is supposedly multilingual, in that one should be able to enter accidentals in German and note names in French, Italian, and Spanish. There are some quirks to this apparent convenience, however. Though Themefinder does not mention it, one can apparently enter "s" instead of "+" for "sharp." This greatly complicates searches that incorporate German note names, as a search for "as" and "ais" yields the exact same result set of themes beginning with A#. This becomes especially confusing when a search for "es," which is equivalent to E-flat in German, produces absolutely no results even though there are 2,469 registered themes that begin with E-flat.

Themefinder's interval search involves several components for known intervals. These components must be entered in a specific order: direction (ascending or descending), quality (major, minor, augmented, diminished, or perfect), and diatonic interval size (1=unison, 10=tenth, etc.). Any of these three compulsories may be left unspecified; in such a case, the search would include any possible combination from that string. So, a search for m6 (minor sixth) would yield results for both ascending and descending intervals with that diatonic size. Searching by scale degree is straightforward and must be entered numerically, from "1" as the tonic to "7" as the subtonic. This search method can only be used for the Classical and Folksong databases in Themefinder. There are no wildcard options for the scale degree search at this time. Melodic contour searching allows users to indicate general and specific ascending and descending intervals, as well as repeated notes. With the refined contour option, users can distinguish between steps and leaps.

Also included in Themefinder is an incomplete, motley index of composers, some of whom have not one theme represented in Themefinder. Each composer name in the index is linked to a works list; works that are cataloged in Themefinder are hyperlinked to the relevant theme. The complete works of both Puccini and Shostakovich are listed, yet there are no themes available by Puccini and only a handful by Shostakovich. As the option exists for users to contribute relevant links to certain themes, users are also encouraged to add composer works lists to the site. Since the submitted data must be formatted in such a way that Themefinder can read it, the user is provided with guidelines explaining this process for three sections. The first, biographical information, is straightforward, if limited, as it includes a composer's name, along with birth and death dates and locations. For more prolific composers, the user is asked to input paging information, which serves to separate a works list's multiple genres from one another, i.e. Bach's cantatas from his motets. If the composition is going to be linked to a theme, the third section—composition information—is perhaps the most crucial of the three divisions. Information entered here will appear in the theme's record.

There is a confusing occurrence regarding the number of search results in some cases. A search within the Classical database for themes with a meter of 4/4 yields 2572 results. The initial results page shows the first ten themes ordered alphabetically by composer, with a menu of hyperlinked subsequent pages numbered up to page 10. There is no option to display more than ten results per page and, most curiously, there is no obvious mechanism that allows the user to browse the 2,472 results past the 100th that appears on page 10. A search with this many results should probably be refined anyway, but there should also be a way to see all results or, if the software is limited to displaying only 100 results per search, Themefinder should provide users with this information.

The order of results seems to vary depending on the database one is searching. Results in the Classical database appear alphabetically by the composer's last name, then alphabetically by work title within each composer. The Renaissance database shows a similar arrangement of results, yet

it adds a level of refinement by displaying a work's multiple voices in order (voice 2 of 5 appears before voice 3 of 5). No obvious hierarchies are displayed in the results lists of either the Folksong database or the RISM collection. When a search is performed across all databases, results from the Classical database are listed first, followed by those from the Essen Folksong database, then themes from the Renaissance database. The RISM incipits appear last in cross-collection search results. Though the ability to sort search results may not be as crucial in known-incipit searching as it is in more traditional databases, the option to sort within large results sets would allow for more refined browsing capabilities in Themefinder.

### RISM Data Within Themefinder

In late 2008, the AMS/MLA Joint RISM Committee voted unanimously to allow the data derived from the US RISM Project at Harvard University to be made freely accessible and searchable within Themefinder. This site is limited to American holdings of music manuscripts and provides access to over 55,000 incipits of music from the seventeenth and eighteenth centuries. The RISM search interface differs from Themefinder's standard search options in that it is much more textually-based. Independent free-entry search boxes are available for composer, title, and keyword. There are drop down menus that offer controlled vocabularies for a theme's genre and the work's holding library.

Currently, it appears that one must access Themefinder's RISM interface through a direct URL, as there is no obvious or even convoluted method of navigating to the RISM database from one of Themefinder's main pages. In addition, it appears that the RISM content displays independently from Themefinder's original three catalogs, as a search for "Bach" and "Clavier" in the RISM incipits yielded sixty-four records titled "Das Wohltemperierte Clavier," while the same collection of works in Themefinder's Classical database shows forty-eight records with titles of the "Well-tempered Clavier." It is worth mentioning, however, that theme records in the Renaissance database display RISM numbers for an incipit's associated work. The RISM incipits display icons similar to the

rest of the themes. If one clicks on the "i" icon, however, one will note that the metadata associated with the RISM themes is much richer than that of a typical Themefinder record. The field numbers in a RISM record may look familiar as they reflect the bibliographic standards adopted by the U.S. RISM committee. The most obvious advantage to the integration of the RISM data into Themefinder is that a user can now perform a much more incipit-focused search. And while the very familiar EBSCO-host interface may be preferable to many users, Themefinder's interface offers many of the same search options in combination with incipit searching. It is unknown to this reviewer whether or not there are plans to include RISM incipits in Themefinder at an international level, but that would seem to be a logical next step.

### Conclusion

Compared to the landmark publication *Dictionary of Musical Themes* (Harold Barlow and Sam Morgenstern [New York: Crown Publishers, 1948]), Themefinder is likely more useful for known-item searching while the *Dictionary* may better facilitate browsing and discovery. Themefinder has, of course, impressively expanded upon Barlow and Morgenstern's original intentions by offering enriched metadata, multiple search options, and enhanced scope. It is obvious to this reviewer that when the CCARH conceived, developed, and released Themefinder, functionality, data entry, and search accuracy were higher priorities than usability. Regardless, the advantageous implications of such a convenient and richly-populated resource in music teaching and research are readily apparent. Perhaps Themefinder's overall usefulness could be expanded if it interacted actively and robustly with other free and reliable resources that supplement its data, such as *NMA Online* [<http://www.nma.at/default-english.htm>], the *Bach Bibliography* [<http://www.npj.com/bach/>], or any number of digital score collections and digitized archives available on the Web. Such a strategy would greatly enrich Themefinder's offerings to the academic community without the strain of revising its own metadata or scope.

ANNE SHELLEY  
University of Minnesota